

SPECIFIC AIMS

While computational techniques are currently widely used in pharmaceutical drug discovery, current generation technologies (such as docking) are unsuitable for true molecular design. Specifically, these techniques fail to predict small molecule binding affinities to target and antitarget biomolecules with sufficient accuracy for the variety of targets currently of interest. While these techniques achieve some success at predicting binding modes, this is not enough. Computational screening techniques do better than random selection of compounds, but they lack the accuracy to guide molecular design or optimization. A new generation of physical techniques, alchemical free energy calculations, are poised to fill this void by providing a quantitative, predictive tool that can be used in multiple stages of the drug discovery pipeline, including lead optimization to drive affinity and selectivity or the retention of potency as other physical properties are optimized.

While alchemical methods have recently seen some spectacular successes, the domain of applicability of these techniques is currently highly limited; to be broadly applicable in accelerating drug discovery, these techniques need further evaluation, refinement, and development before they are ready for routine, diverse applications. There is a vast gulf between targets within the domain of applicability and those which are outside it. To rapidly expand the domain of applicability and improve these techniques, we need carefully selected systems of intermediate complexity to bridge this gulf. Without it, these techniques may encounter the same problems faced by docking and related techniques: routine failure without clear insights into why, and years to decades spent in making small methodological modifications without dramatic improvements in predictive power. We propose to identify and develop a set of model systems that span the gulf between pharmaceutical targets inside and outside the current domain of applicability, generating new high-quality experimental data and running blind challenges design to push the field forward rapidly. The generated data – and the resulting challenges – will result in a dramatically increased domain of applicability and will push these techniques into standard application in drug design. In addition to generating this experimental data, we will conduct reference calculations of the values measured, making these, and the associated input files and software tools, available to the community as a point of comparison to further facilitate advances.

Here, we seek to rapidly drive development of computational methods for drug discovery by generating targeted data for a series of blind SAMPL (“Statistical Assessment of Modeling of Proteins and Ligands”) challenges. Specifically, we will advance these techniques via the following efforts:

Aim 1. Extend SAMPL physical property challenges.

We will build on the enormous success of previous SAMPL hydration free energy and distribution coefficient challenges by developing new solution-phase datasets for druglike small molecules. These challenges both test critical aspects of small molecule modeling (including accounting for interactions and treatment of protonation/tautomeric state) and improve the ability to predict physical properties relevant to drug discovery in regions of new chemical space. We will initially focus on distribution between organic phases and on pKa's and their modulation by solvent environment, using this data to drive improvements in the modeling of ligand interactions.

Aim 2. Broaden introductory SAMPL ligand binding challenges.

We will cultivate new host-guest binding data to serve to span between physical property prediction and protein-ligand binding. Host guest systems are some of the simplest cases of molecular recognition, and thus host-guest binding data will drive improvements in the modeling of simple binding systems with techniques of relevance to drug discovery.

Aim 3. Generate advanced model systems for protein-ligand binding with biological relevance. We will identify and cultivate and cultivating suitable protein targets (determined by being difficult but tractable for physical techniques in order to push their limits) as model ligand binding systems, collecting high quality binding data, running blind challenges on these datasets, and rapidly releasing the data to the community in an ongoing, rolling fashion.

This study will also involve an extensive computational component as part of each of these aims, as we will calculate reference values with a standard force field for the properties (partitioning, distribution, or binding) being measured.

Overall, this work will drive a series of cycles of testing and innovation for physical methods for predicting binding and physical properties, providing a foundation for the next several generations of computational methods for pharmaceutical drug discovery.